2873

METEROLOGIC DATABASE

DORSET RESEARCH CENTRE

MARCH 1994

Ontario

Ministry of
Environment
and Energy

ISBN 0-7778-1923-6

METEOROLOGIC DATABASE

DORSET RESEARCH CENTRE

MARCH 1994

Cette publication technique
n'est disponible qu'en anglais.

Copyright: Queen's Printer for Ontario, 1994
This publication may be reproduced for non-commercial purposes
with appropriate attribution.

PIBS 2873

METEOROLOGIC DATABASE

DORSET RESEARCH CENTRE

Report prepared by:

B.A. Hutchinson
L.D. Scott
M.N. Futter
A. Morgan
W. Coté

Dorset Research Centre
P.O. Box 39, Bellwood Acres Road
Dorset, ON P0A 1E0

TABLE OF CONTENTS

1. Introduction
2. Database Design
3. Table Structures

3.1 Data Dictionary

M_COLUMN_DESCRIPTION	Database Dictionary Table

3.2 Data Tables

M_DAILY	Daily meteorologic observations
M_HOURLY	Hourly meteorologic observations

3.3 Description Tables

M_METHOD	Meteorologic parameter method codes by station
M_METHOD_DESCRIPTION	Meteorologic method descriptions
M_QC	Quality control flags and their descriptions
M_STATION	Station spatial locations
M-PARAM_DAILY	Meteorologic parameters recorded daily
M_PARAM_HOURLY	Meteorologic parameters recorded hourly

3.4 Loading and Test Tables

M_UPDATE_DAILY	Data updated from M_DAILY
M_UPDATE_HOURLY	Data updated from M_HOURLY
M_UPDATE_SESSION	Information about update sessions
M_DAILY_TEMP_LOAD	Temporary loading table for M_DAILY

M_HOURLY_TEMP_LOAD	Temporary loading table for M_HOURLY
M_DAILY_TEMP_TEST	Temporary testing table for M_DAILY
M_HOURLY_TEMP_TEST	Temporary testing table for M_HOURLY

4. Database Summary Listing

5. Database Loading and Update Procedures

5.1 Database Loading
DRC Data
AES Data

5.2 Database Back-up Strategy

6. References

Appendix 1 - Example SQL Retrievals

LIST OF TABLES

Table 1 Chart showing cross-references between meteorologic data tables by key column names.

Table 2 Current contents of the quality control reference table (M_QC) as of February, 1993.

Table 3 Station number and abbreviation information for DRC stations.

LIST OF FIGURES

Figure 1 Map of study area meteorologic monitoring sites.

Figure 2 Data flow between the daily meteorologic database (M_DAILY) and some of the other DRC tables.

Figure 3 Update procedures (including testing and loading) for the DRC meteorologic database.

1. INTRODUCTION

The meteorologic database at the Dorset Research Centre (DRC) consists of daily and hourly observations along with descriptions of the sites and monitoring equipment. These monitoring sites are located mainly in the Muskoka-Haliburton district of southern Ontario although there are several associated sites outside this district (see Figure 1). The data are collected by several agencies including Atmospheric Environment Services (AES), Air Resources Branch of MOEE (ARB), Canada Centre for Inland Waters (CCIW), Ontario Ministry of Natural Resources (MNR) and the DRC (Limnology Section, Water Resources Branch). Data collection methodology reports are available from these agencies (Locke and Scott 1986, Locke and deGrosbois 1986 and MANOBS 1991). Much of the research performed at Dorset involves modelling of whole watershed processes. Both meteorology and atmospheric deposition measurements are vital to the development and validation of these models. Figure 2 illustrates some of these data processes with links to the daily meteorologic database. This meteorologic database is available to many groups involved in co-operative studies with the DRC.

This document describes the structure and location of the tables in the meteorologic database. A description of each parameter stored in the database and ancillary parameters collected are also given. Procedures used to load the "raw" data to the meteorologic database as well as quality assurance and control methodologies are discussed. Several in-house programmes have been developed to manage these data. The quality control procedures are described in Hutchinson et al. (1993) along with update and testing procedures. These procedures are documented as performed in a database log book.

2. DATABASE DESIGN

All meteorologic data are stored in an ORACLE relational database (ORACLE Corp. 1991). Table design and labelling conventions conform to standards set by the DRC for all ORACLE databases maintained by the Limnology Section at Dorset (LaZerte et al. 1990,

Figure 1. Map of Study Area Meteorologic Monitoring Sites.

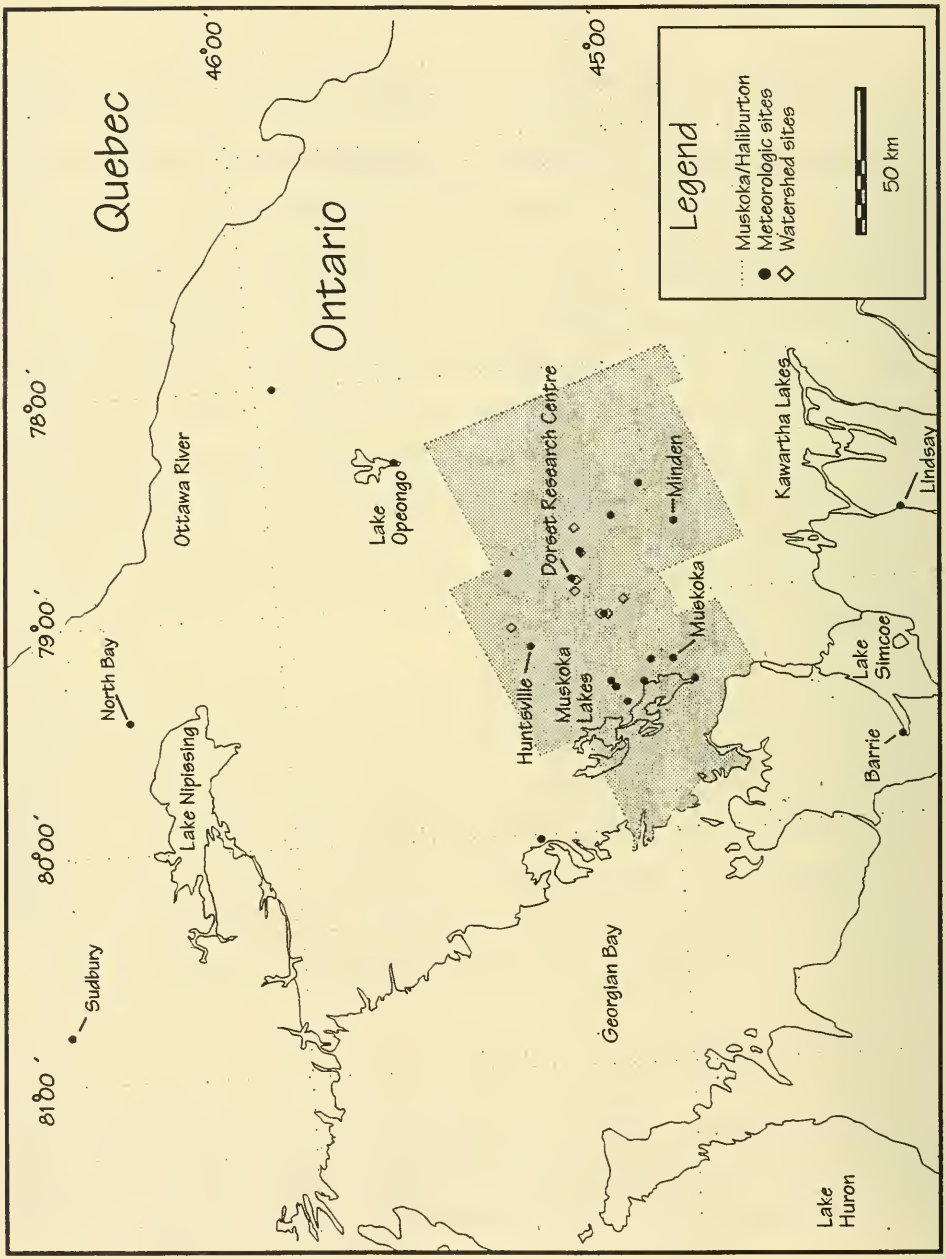
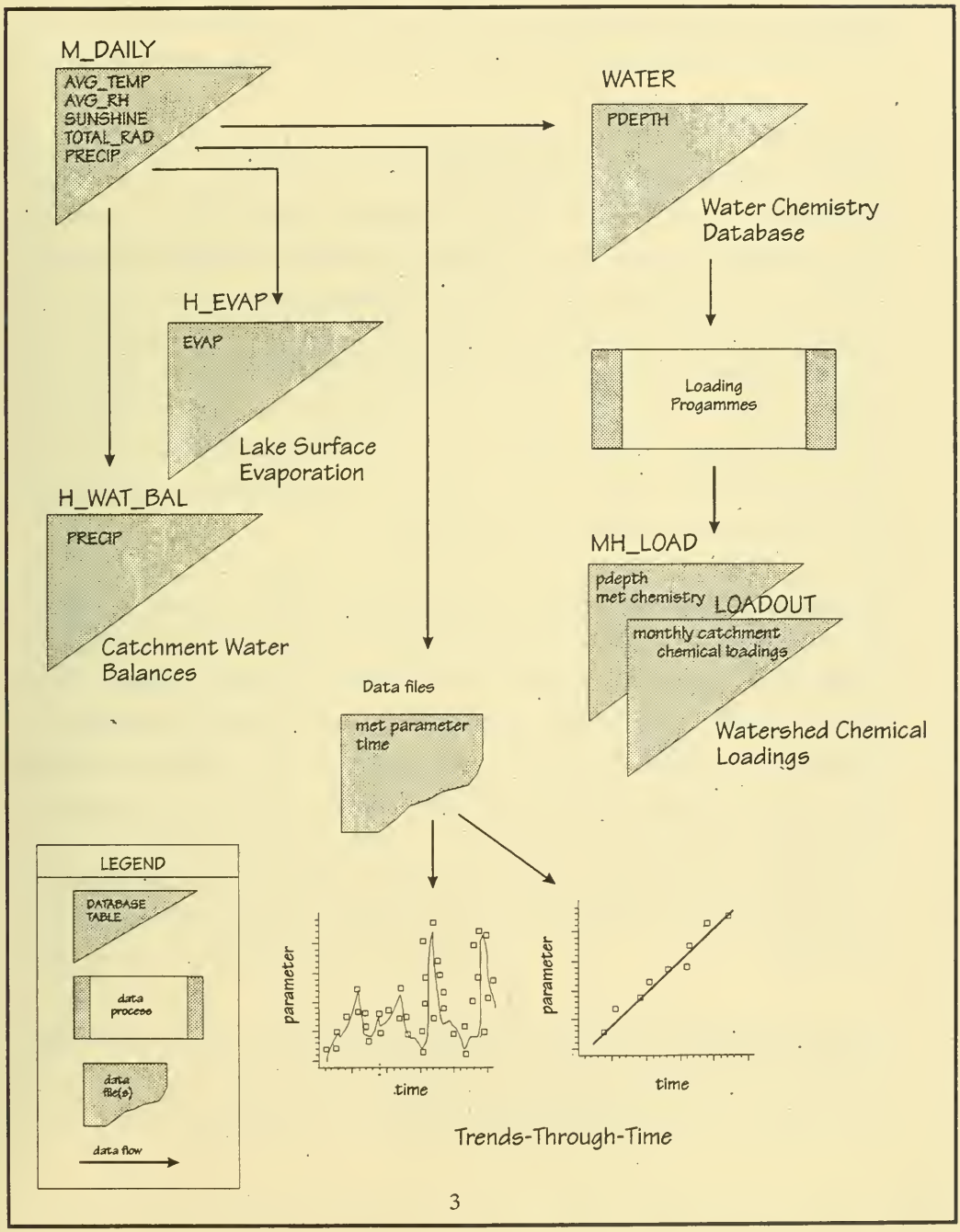


Figure 2. Data Flow Bewteen the Daily Meteorologic Database (M_DAILY) and some of the other DRC Databases.



Futter et al. 1993 and Pawson and Yan 1993). The database resides on an HP-9000 mini computer system running HP-UX and is directly accessible from UNIX terminals, as well as remotely through DOS and OS/2 work-stations attached to the DRC thin Ethernet LAN.

Data is collected by the DRC using Campbell Scientific 21X data loggers. The data is retrieved weekly from these data loggers by down-loading the current data to a solid state storage module. These devices are read at a dedicated PC using software developed by Campbell Scientific (Campbell Scientific 1989). The field data is transferred to an optical disk for long term archiving once every two weeks. After preliminary editing and re-formatting of the hourly and daily data, the files are transferred to the HP-9000 mini-computer. The procedures to transfer data from the field to the ORACLE database on the HP-9000 are fully documented in MOEE internal procedures manuals (Hutchinson et al. 1993). AES and ARB use data loggers for some of the measured parameters at airport locations (MUS, NBY, SUD). All other data are recorded by field crews or volunteer observers using manual instrumentation (MANOBS, 1991).

The database may be queried using any one of the standard access methods. These include SQL*Plus, ORACLE for 123 or Q (Lazerte 1990). The database may be accessed across the LAN directly through SQL*TCP/IP (ORACLE 1990a), or through an OS/2 gateway using SQL*NMP (ORACLE 1990b). Several data request queries have also been developed to provide standard input for DRC application programmes requiring meteorologic data. These queries are either hard coded in the application programmes, or are contained in UNIX script files stored in system support directors. Frequent queries to the database often entail statistical summaries of the data over a specified time period. Appendix 1 illustrates two sample queries that return a single row of data of a multi-row and column query to a database table.

All tables in the database are stored in the "meteor" table space. Indices for these tables are stored in the "meteor_index_space" table space. All tables are owned by the "met_data_mgr" account. While the data is publicly available, updates and modifications to

the database can only be performed by the database manager, or other staff as authorized by the database manager. Programmes used for modifying the database are described in Section 5.1 of this document.

Data are collected on three different schedules. DRC data is collected on a daily and hourly basis and AES data is collected morning and afternoon. Both daily and hourly data are collected at all DRC sites. Hourly data is reported at the end of the hour for which it was recorded. Daily data is recorded on the basis of a calendar day at the DRC sites (from midnight to midnight). Federal sites (AES) follow a different daily protocol. AES daily data is summarized between 0800 of calendar day one to 0800 of calendar two. This results in an offset of eight hours and subsequent differences between the values collected by AES and DRC.

3. TABLE STRUCTURES

The storage of the meteorologic database has been divided into data, description and update categories. Each of the tables have been designed with key columns that will link to many other tables. Table 1 illustrates the cross references that exist between the database tables. The meteorologic database can be linked to other databases stored at the DRC including hydrology and water chemistry. References to LaZerte et al. (1990) and Futter et al. (1993) show how these links may be made (especially using the STN and SDATE keys). The table names are contained in the top row and first column of Table 1. Where tables can be keyed together directly, the cell where the row and column intersect contains the key column name in the two tables. It is not possible to directly link tables for which the intersection cell of the row column combinations is blank, however, these tables may be linked through a third table with key connections to the first and second.

A quality control flag, or QC code, is associated with each observed parameter in the suite of daily and hourly meteorologic observations. This effectively doubles the size of the database since for each data column, there is a corresponding QC column. The daily and

hourly databases generally contain only real measured values. In some cases, it is possible to use other methods (such as recording strip charts) to fill in missing values. A QC code is used to explain why the data field is NULL.

Table 1 This table shows the table names in the top row and first column. Where tables can be keyed together directly, the cell where the row and column intersect contain the key parameters in the two tables. It is not possible to directly key tables for which the intersection cell of the row column combinations is blank.

	M_DAILY	M_HOURLY	M_UPDATE_HOURLY	M_UPDATE_DAILY	M_UPDATE_SESSION	M_METHOD	M_QC	M_PARAM_DAILY	M_PARAM_HOURLY	M_STATION
M_DAILY		STN SDATE		STN SDATE PARAM.		STN SDATE PARAM.	Q_C #	STN PARAM		STN
M_HOURLY	STN SDATE		STN SDATETIME PARAM.			STN SDATETIME PARAM.	Q_C #		STN PARAM.	STN
M_UPDATE_HOURLY		STN SDATETIME PARAM.			SKEY					
M_UPDATE_DAILY	STN SDATE PARAM.				SKEY					
M_UPDATE_SESSION			SKEY	SKEY						
M_METHOD	STN PARAM. SDATE	STN PARAM. SDATETIME								
M_QC	Q_C #	Q_C #								
M_PARAM_DAILY	STN PARAM.									STN
M_PARAM_HOURLY		STN PARAM.								STN
M_STATION	STN	STN						STN	STN	

3.1 Data Dictionary

M_COLUMN_DESCRIPTION

The M_COLUMN_DESCRIPTION table provides the an abbreviated version of the column definitions contained in this document in an on-line format. There is one row in this table for every column of the tables in Sections 3.2 through 3.4

Column	Data Type	Example
TABLE_NAME	char(30)	M_DAILY
COLUMN_NAME	char(30)	TOTAL_RAD
COMMENTS	char(255)	The total daily unfiltered radiation for SDATE, monitored by DRC sites only.

TABLE_NAME The name of the table.

COLUMN_NAME The name of the column within the table.

COMMENTS An abbreviated version of the column description contained in this document.

3.2 Data Tables

M_DAILY Table

This table contains one daily data value per calendar date for each parameter collected at the monitoring sites. Descriptions of parameters monitored for each station are documented in the M_PARAM_DAILY table. Two indexes, M_DAILY_UNIQ_INDEX and M_DAILY_INDEX, have been created on the table M_DAILY. M_DAILY_UNIQ_INDEX contains a unique composite key for each record in the M_DAILY table on the columns STN and SDATE. This index prevents duplication of records in the database. The index M_DAILY_INDEX includes all parameters collected at the DRC monitoring sites. This index facilitates faster data retrieval.

Column	Data Type		Example***	Units	Agency
*STN	not null	char(4)			
*SDATE	not null	date		DD-Mon-YY	
STATUS	null	char(1)			
AVG_TEMP	null	number	-8.2	°C	AES, DRC
MAX_TEMP	null	number	23.7	°C	AES, DRC
MIN_TEMP	null	number	-15.9	°C	AES, DRC
PRECIP	null	number	3.2	mm	AES, DRC
DP_TEMP	null	number	8.4	°C	AES
AVG_RH	null	number	76.5	%	AES, DRC
MAX_RH	null	number	91.1	%	AES, DRC
MIN_RH	null	number	72.3	%	AES, DRC
SUNSHINE	null	number	0.9	hrs	AES
PAR_RAD	null	number	324.2	kJ/m^2	DRC, CCIW
TOTAL_RAD	null	number	678.9	kJ/m^2	DRC
VAPOUR_PRESS	null	number	1.23	Pa	DRC
AVG_ATMOS	null	number	34.9	KPa	AES, DRC
MAX_ATMOS	null	number	36.8	KPa	DRC
MIN_ATMOS	null	number	33.0	KPa	DRC
RAIN	null	number	3.2	mm	AES
SNOW	null	number	3.2	cm	AES
GRND_SNOW	null	number	36	cm	AES

Column	Data Type		Example***	Units	Agency
BLFT_WT	null	number	15.67	weight (mm)	DRC
BLFT_MV	null	number	0.102	mV	DRC
BLFT_STD	null	number	1.804	mV std. dev.	DRC
WIND_SPEED	null	number	2.3	ms^{-1}	AES, DRC
WIND_DIR	null	number	275	deg	DRC
WIND_MAG	null	number	1.971	ms^{-1}	DRC
WIND_STD	null	number	23.68	deg std. dev.	DRC
SO_2	null	number	0.17	ppm	ARB
NOX	null	number	0.86	ppm	ARB
SO_4	null	number		ug/m3	ARB
NO_3	null	number		ug/m3	ARB
***PARAM_*	null	number	77	missing	AES, DRC

* The composite primary key is derived from STN and SDATE.
** QC code associated with each parameter. QC codes are described in M_QC.
*** Examples shown here illustrate sample data for the Muskoka/Haliburton area.

STN The unique character station abbreviation code assigned by the DRC to each monitoring site.

SDATE The date on which the data were collected.

STATUS Will contain one of two values, NULL or "M". An "M" indicates that the set of observed values for that station on that day have been modified. Changes to any value in the table are recorded in the M_DAILY_UPDATE table. The update procedure is documented in Hutchinson et al. (1993).

AVG_TEMP Mean daily temperature recorded on SDATE. This parameter can be measured in two ways. For the DRC stations, the average temperature is recorded over the twenty-four hour period and is derived from

readings taken at five second intervals. It is not the average of MIN_TEMP and MAX_TEMP. For the AES stations, AVG_TEMP is the average of the two temperature readings taken at 0800 and 1600 hours.

MAX_TEMP The maximum temperature recorded on SDATE.

MIN_TEMP The minimum temperature recorded on SDATE.

PRECIP The total amount of precipitation, recorded as millimetres of water, which was deposited on SDATE. For DRC stations, this is midnight to midnight, at AES stations, the value is reported as the sum of precipitation (1600-0800) the first day, and (0800-1600) the following day. AES stations also record this PRECIP daily total as either rain or snow (or some of both).

DP_TEMP The dew point temperature collected at AES recording sites. Dew point is determined from a wet and dry bulb psychrometer. AES records mean hourly values, which have been converted to a mean daily value and entered to the database. This parameter can also be calculated using an empirical relationship between temperature and average relative humidity (Hutchinson et al. 1993).

AVG_RH The percent average relative humidity recorded for SDATE. The AES datum is collected with a wet and dry bulb psychrometer (MANOBS 1991). At the DRC sites, average relative humidity is the average of readings taken at five second intervals from a relative humidity thermistor probe.

MAX_RH The maximum daily observed relative humidity from the relative

humidity thermistor at STN on SDATE.

MIN_RH The minimum daily observed relative humidity from the relative humidity thermistor at STN on SDATE.

SUNSHINE The total number of hours of bright sunshine at STN on SDATE as determined from a burn pyranometer chart monitored by AES stations located at airports only.

PAR_RAD The daily total photosynthetically available fraction of TOTAL_RAD monitored by DRC sites only.

TOTAL_RAD The total daily unfiltered radiation for SDATE monitored by DRC sites only.

VAPOUR_PRESS The average daily saturation vapour pressure. Saturation vapour pressure is a polynomial function of temperature. The saturation vapour pressure is calculated every five seconds at the data logger. VAPOUR_PRESS is the average of these five second values collected on SDATE. This parameter is monitored by DRC sites only.

AVG_ATMOS The average daily atmospheric pressure on SDATE. At the DRC monitoring sites, this parameter is measured with an atmospheric pressure thermistor. At the AES sites, this parameter is measured with a barometer.

MAX_ATMOS The maximum daily atmospheric pressure recorded at STN on SDATE.

MIN_ATMOS The minimum daily atmospheric pressure recorded at STN on SDATE.

RAIN The daily total recorded depth of rain, in millimeters, on SDATE at STN, by AES stations only.

SNOW The daily total recorded depth of snow, in centimeters, which fell at STN on SDATE, by AES stations only.

GRND_SNOW The total depth of the snow pack on a given date. A zero in the database denotes no snow pack, while a missing value indicates only that no data on the snow pack depth was available.

BLFT_WT The weight of the Belfort precipitation collector at STN on SDATE. BLFT_WT is recorded as the equivalent distance in millimetres by which the precipitation collector bucket is depressed, by DRC stations only.

BLFT_MV The potential, in millivolts, across the potentiometer on the Belfort gauge. BLFT_WT is a linear function of BLFT_MV, by DRC stations only.

BLFT_STD The standard deviation of the BLFT_MV readings for the Belfort precipitation collector at STN on SDATE, by DRC stations only.

WIND_SPEED The average wind speed in metres per second recorded at STN on SDATE.

WIND_DIR The average wind direction, in degrees, at STN on SDATE.

WIND_MAG The magnitude of the vector defined by WIND_SPEED and WIND_DIR at STN on SDATE.

WIND_STD The standard deviation of the wind direction in degrees at STN on SDATE.

SO_2 The amount of sulphur dioxide in the atmosphere. This parameter is collected by ARB staff at Dorset.

NOX The amount of nitrous oxides in the atmosphere. This parameter is collected by ARB staff at Dorset.

SO_4 The amount of sulphate in the atmosphere. This parameter is collected by the ARB staff at Dorset.

NO_3 The amount of nitrate in the atmosphere. This parameter is collected by ARB staff at Dorset.

PARAM_ Every data point in the database has an associated quality control (QC) code. These codes allow for creation of data sets containing only real data (i.e., blank spaces will be left where no data was available) and continuous data vectors where data is filled in or estimated for days on which no reading was available for a parameter at a station. The column labelling convention of these QC codes conforms to other DRC database tables. The data column name is used followed by an underscore to denote QC information. As an example, the QC code column for AVG_TEMP is AVG_TEMP_, and for PRECIP is PRECIP_. Further discussion of quality control codes is provided in Section 3.3 of this document on the M_QC table.

M_HOURLY Table

This table contains 24 hourly rows per calendar date for each monitored site. Descriptions of meteorologic parameters monitored for each station are provided in the M_PARAM_DAILY table.

Column	Data Type		Example***	Units	Agency
*STN	not null	char(4)			
*SDATETIME	not null	date.time	01-JAN-89.0100	DD-Mon-YY.HH24Mi	
SDATE	not null	date	01-JAN-89	DD-Mon-YY	
STATUS	null	char(1)			
AVG_TEMP	null	number	-8.2	°C	DRC
PRECIP	null	number	3.2	mm	DRC
DP_TEMP	null	number	8.4	°C	AES
AVG_RH	null	number	76.5	%	DRC
SUNSHINE	null	number	0.9	hrs	AES
PAR_RAD	null	number	564.2	kJ/m^2	DRC
TOTAL_RAD	null	number	678.9	KJ/m^2	DRC
VAPOUR_PRESS	null	number	1.23	Pa	DRC
AVG_ATMOS	null	number	34.9	kPa	DRC
BLFT_WT	null	number	15.67	weight (mm)	DRC
BLFT_MV	null	number	0.102	mV	DRC
BLFT_STD	null	number	1.804	mV std. dev.	DRC
WIND_SPEED	null	number	2.3	ms^{-1}	DRC
WIND_DIR	null	number	275	deg	DRC
WIND_MAG	null	number	1.971	ms^{-1}	DRC
WIND_STD	null	number	23.68	deg std. dev.	DRC
SO_2	null	number	0.029	ppm	ARB
NOX	null	number	0.035	ppm	ARB
SO_4	null	number		ug/m3	ARB
NO_3	null	number		ug/m3	ARB
** *PARAM_*	null	number	84 (trace)		AES, DRC

* The composite primary key is derived from STN and SDATETIME .

** QC codes associated with each parameter.

*** Example shown is representative of the area illustrating order of magnitude for the values.

Two indexes, M_HOURLY_UNIQ_INDEX and M_HOURLY_INDEX, have been created on the table M_HOURLY. M_HOURLY_UNIQ_INDEX contains a unique composite key for each record in the M_HOURLY table on the columns STN, and SDATETIME. This index will prevent duplication of records in the M_HOURLY table. M_HOURLY_INDEX includes all parameters collected at the DRC monitoring sites. This index speeds access during data retrieval.

STN The character abbreviation of the station name assigned by Limnology Section staff.

SDATETIME The date and time of the sample. Twenty four hourly samples are collected each day. Sample times are the end of the sampling period. For example, 1400 records the values observed between 1300.01 and 1400 hours. Due to ORACLE database constraints, samples from 2300.01 to 2400 are assigned a time code of 2359.

SDATE The date on which the sample was collected. This parameter (with STN) can be used for keying to other tables.

STATUS Contains a code describing whether parameters collected at STN on SDATETIME have been modified. A NULL value indicates that no modifications have taken place. An "M" indicates that at least one parameter collected at STN on SDATETIME has been modified. The changed parameter and its old value are stored in M_HOURLY_UPDATE.

AVG_TEMP The average temperature recorded at STN on the hour ending at SDATETIME.

PRECIP The amount of precipitation, expressed as millimetres of water, which fell at STN at SDATETIME. Each hourly value represents an average of 12 readings taken per minute for the last 5 minutes of the hour (60 value).

DP_TEMP The dew point temperature recorded in degrees Celsius at STN at SDATETIME.

AVG_RH The percent average relative humidity recorded at STN at SDATETIME.

SUNSHINE The fraction of the hour for which there was bright sunshine at STN at SDATETIME.

PAR_RAD The photosynthetically available radiation at STN at SDATETIME.

TOTAL_RAD The total unfiltered radiation for STN at SDATETIME.

VAPOUR_PRESS The average saturation vapour pressure calculated for STN at SDATETIME. Saturation vapour pressure is a polynomial function of temperature.

AVG_ATMOS The average atmospheric pressure as recorded by an atmospheric pressure thermistor at STN at SDATETIME.

BLFT_WT The distance, in millimetres, by which the Belfort precipitation recording system is depressed. This represents the amount of precipitation (as a weight calibrated in mm).

BLFT_MV The potential difference, in millivolts, across the Belfort recording system. BLFT_WT is a linear function of BLFT_MV.

WIND_SPEED The average wind speed at STN at SDATETIME measured in meters per second.

WIND_DIR The average wind direction at STN at SDATETIME measured in degrees.

WIND_MAG The magnitude of the wind speed wind direction vector at STN at SDATETIME.

WIND_STD The standard deviation associated with the average wind direction for STN at SDATETIME.

SO_2 The amount of sulphur dioxide present in the atmosphere at STN at SDATETIME.

NOX The amount of nitrous oxides in the atmosphere at STN at SDATETIME.

SO_4 The amount of sulphate in the atmosphere at STN at SDATETIME.

NO_3 The amount of nitrates in the atmosphere at STN at SDATETIME.

PARAM_ The quality control code associated with each parameter. For example, PRECIP_ is the QC code for PRECIP. A QC code is available for each parameter observation with information describing methods used to estimate any missing or erroneous data.

3.3 Description Tables

M_METHOD Table

M_METHOD contains information on the methods employed in data collection at the DRC monitoring sites. The METHOD_ID column points to the M_METHOD_DESCRIPTION table for complete descriptions of the instrumentation used. Currently active equipment will have a NULL value in the END_DATE field. The table has a composite primary key of STN, START_DATE and METHOD_ID.

Column	Data Type		Example**
*STN	not null	char(4)	
*START_DATE	not null	date	01-Jun-89
END_DATE	null	date	31-May-91
PARAMETER	not null	char(10)	AVG_TEMP
METHOD_ID	null	number(4)	method identification code
DAILY_HOURLY	null	char(1)	daily/hourly flag

* The composite primary key is derived from STN and START_DATE.
** Descriptions of parameters monitored for each station are provided in M_PARAM_HOURLY and M_PARAM_DAILY.

STN The unique character abbreviation of the station name assigned by Limnology Section staff.

START_DATE and END_DATE
The date range over which the sample method was/is in use.

PARAMETER The column name assigned for each meterologic parameter recorded in the database tables M_DAILY and M_HOURLY.

METHOD_ID A numeric code used to describe the methods used for each meteorologic paramter.

DAILY_HOURLY A flag to denote whether the parameter is observed on a daily or hourly basis.

M_METHOD_DESCRIPTION Table

This table contains detailed information about each meteorologic parameter. It points back to the M_METHOD table via the METHOD_ID column.

Column	Data Type		Example**
METHOD_ID	not null	number	11
INSTRUMENT	null	char(40)	Type 45 Anemometer
METHOD_DESC	null	char(80)	Wind Speed
AGENCY	null	char(4)	AES
UNITS	null	char(10)	m/sec
PRECISION	null	char(10)	0.1 m/sec

METHOD_ID A numeric code used to describe the methods used for each meteorologic parameter.

INSTRUMENT A description field of the instrumentation used by this method for a parameter.

METHOD_DESC A description field of the sampling method for a parameter.

AGENCY The agency recording the meteorologic observations.

UNITS The units the parameter is reported in.

PRECISION The precision associated with the parameter.

M_QC Table

This table contains the quality control information used as values in the QC columns for each parameter collected in M_DAILY and M_HOURLY. Table 2 contains the current contents of this table (February 1993).

Column	Data Type	Type	Example
Q_C	not null	number	84
REMARK	not null	char(50)	Trace. Value is zero

Q_C A number representative of a quality control flag assigned for a data value. Some of the numbers have been derived from the ASCII ordinate of a single letter used in the QC remark description (e.g., 77 = "M" for Missing values).

REMARK A description field to explain the application of the QC code.

Table 2 Current contents of the M_QC table (February, 1993)

Q_C	Remark	Agency
32	Valid data, QC column (i.e., PRECIP_) value will be blank	DRC
65	Accumulated amount: previous value was zero	AES
66	More than one occurrence and estimated	AES
67	Precipitation occurred, amount uncertain, value is zero	AES
68	Derived	AES
69	Estimated	AES
70	Accumulated and estimated	AES
71	Uncorrected	AES
72	Freezing	AES
73	Unadjusted	AES
77	Missing	AES,DRC
78	Null	AES,DRC
80	Potentiometer was failing (86/11/01-88/01/31)	DRC
84	Trace. Value is zero	AES
99	Estimated from Chart recordings	DRC
101	Erroneous	DRC

M_STATION Table

This table contains information on the location of stations for which data is available in the meteorologic databases.

Column	Data Type		Example**
*STN	not null	char(7)	DWI
*STATION	not null	char(11)	6082178
START_DATE	not null	date	01-Jan-76
END_DATE		date	
INFO		char(40)	Dwight
DISTRICT		char(15)	Muskoka
TOWNSHIP		char(20)	Ridout
LATITUDE		number	45 13
LONGITUDE		number	78 56
EASTING		number	5009675
NORTHING		number	662375
ELEVATION		number	323
WATERSHED		char(6)	2EB
ZONE		number	17
TOP_MAP		char(6)	31E2
AGENCY		char(4)	AES

* The primary key is derived from STN.
** Descriptions of parameters monitored for each station are provided in M_PARAM_HOURLY and M_PARAM_DAILY.

STN The unique character station abbreviation code assigned by the Limnology Section.

STATION An eleven digit key to WATER.STATION (the station code in the WATER table) for monitoring sites maintained by the Dorset

Research Centre. Those sites not maintained by the Dorset Research Centre, use the station code assigned by the monitoring agency.

START_DATE The earliest date for which there is data in the Limnology Section databases for STN.

END_DATE Either the date at which the station was decommissioned, or the most recent date for which there is data for that station in the permanent meteorology database.

INFO A character field containing additional information about the sample collector site.

DISTRICT The District (or County) where the STN is located.

TOWNSHIP The Township where the STN is located.

LATITUDE and LONGITUDE

Are the latitude and longitude coordinates of the monitoring station, expressed as degrees minutes seconds.

EASTING and NORTHING

Are the Universal Transverse Mercator (UTM) coordinates for the collecting station.

ELEVATION The height of the monitoring station, expressed in feet, above sea level.

WATERSHED The code identifying the tertiary watershed in which the monitoring site is located.

ZONE	The UTM zone in which the STN is located.
TOP_MAP	The 1:50,000 topographic map where the majority of the site is located.
AGENCY	The abbreviation for the agency collecting the data. Currently, the database contains data from four agencies: AES (Atmospheric Environment Services, Canada), ARB (Air Resources Branch, Ontario Ministry of Environment & Energy), DRC (Dorset Research Centre) and MNR (Ontario Ministry of Natural Resources).

M_PARAM_DAILY and M_PARAM_HOURLY Tables

The M_PARAM_DAILY and M_PARAM_HOURLY tables are designed to simplify the process of determining which parameters are measured at each site in either of the meteorologic databases. The tables are set up such that a user can rapidly determine those parameters collected at each station by submitting a query of the form:

```
SELECT * FROM M_PARAM_DAILY
WHERE STN = "station abbreviation";
```

or

```
SELECT * FROM M_PARAM_HOURLY
WHERE STN = "station abbreviation";
```

Column	Data Type	Example
*STN	not null	CHAR(4)
PARAMETER	not null	CHAR(15)
PARAM_DESC	null	CHAR(240)

* The primary key is derived from STN.

STN The unique character station abbreviation code assigned by the Limnology Section.

PARAMETER The abbreviated name used in reference to parameter stored in the database.

PARAM_DESC Additional descriptive information about the parameter.

3.4 Loading and Test Tables

The loading of new data to the main database (M_HOURLY and M_DAILY) and the quality assurance protocols follows a well defined series of steps. These steps involve the flow of data through a sequence of database tables, processed by DRC developed application programmes. Figure 3 illustrates this data flow process. The loading, testing and update tables are described in this Section.

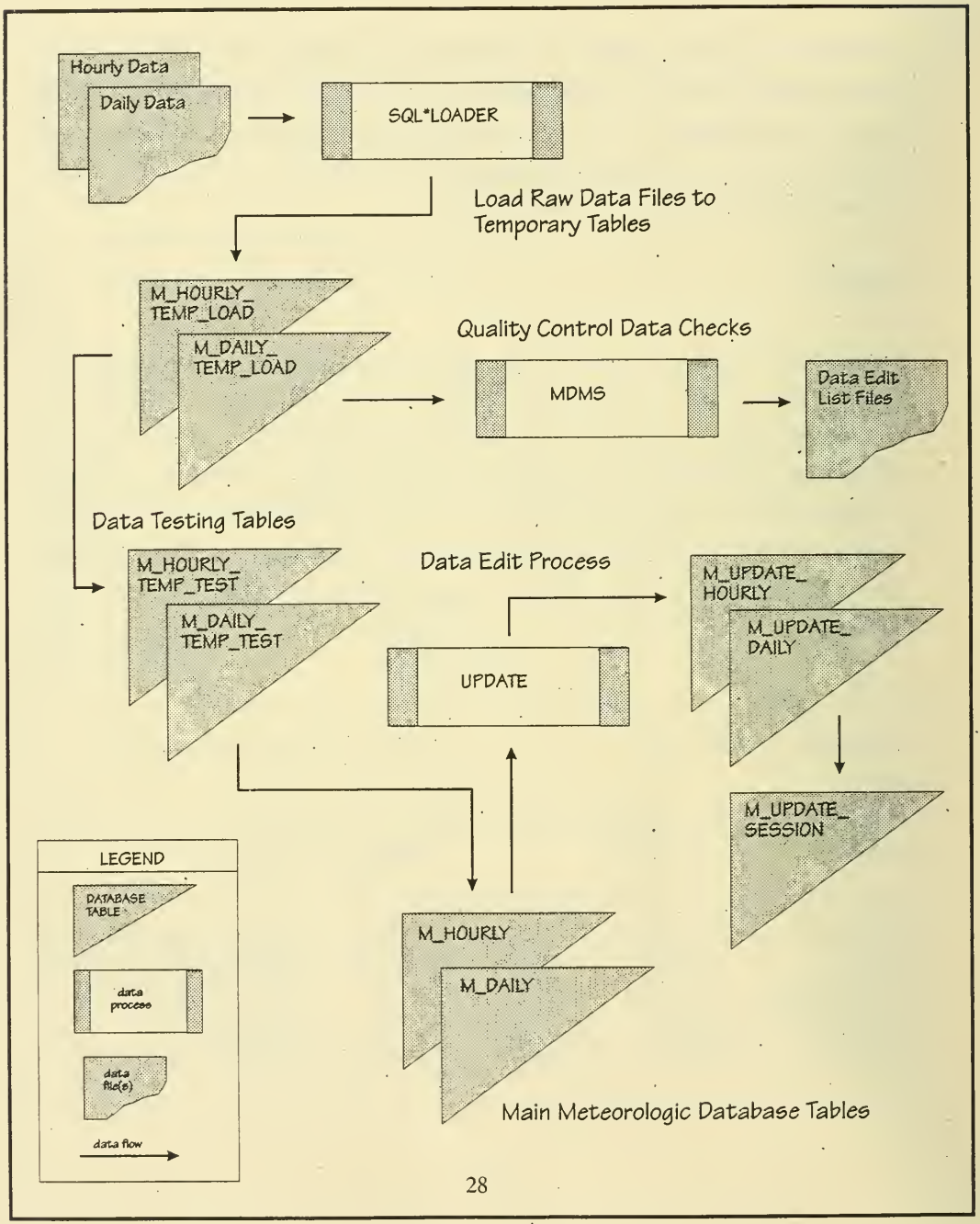
UPDATE

Three tables are used to document changes to the data in M_DAILY and M_HOURLY. These tables are M_UPDATE_SESSION, M_UPDATE_HOURLY and M_UPDATE_DAILY. M_UPDATE_SESSION stores information such as the date of the update, reason for the update, and the operator who performed it. When a change is made to a record in either the M_HOURLY or M_DAILY table, the status flag in the "updated" table is set to "M" (modified). The station, date, parameter which was changed, its old value, and associated QC code (if any) are stored in either the M_UPDATE_DAILY or M_UPDATE_HOURLY table.

M_UPDATE_DAILY Table

Column	Data Type		Example	Description
STN	not null	char(4)	DOR	station being updated
SDATE	not null	date	01-Jul-90	date of record being updated
SKEY	not null	number	22	key to update session
PARAMETER	not null	char(10)	RAIN	parameter updated
RESULT	null	number	3.2	old result
RESULT_	null	number	67	QC code

Figure 3. Update Procedures (including testing and loading) for the DRC Meteorologic Database.



M_UPDATE_HOURLY Table

Column	Data Type		Example	Description
STN	not null	char(4)	DOR	station where record has been updated
SDATETIME	not null	date	01-May-87.1400	date and time of record which was updated
SKEY	not null	number	22	key to update session
PARAMETER	not null	char(10)	RAIN	parameter updated
RESULT	null	number	3.2	old result
RESULT_	null	number	67	QC code

M_UPDATE_SESSION Table

Column	Data Type		Example	Description
SKEY	not null	number	22	key to update session
UDATETIME	not null	date	01-May-91.1400	dd-mon-yy
PERSON	null	char(10)	BAH	person who changed the value

TEMPORARY LOADING Tables

Temporary loading tables are used to gather data in ORACLE format as they are retrieved from field collection. Initial editing checks are performed on the table data as they are entered and after approximately one months worth of data has accumulated in the tables, all records are moved into temporary testing tables where they are checked prior to being inserted into the permanent tables.

The two temporary loading tables are M_HOURLY_TEMP_LOAD and M_DAILY_TEMP_LOAD. The structure of these two tables is identical to that of M_HOURLY and M_DAILY.

TEMPORARY TEST Tables

After the data have been loaded and have passed initial visual checks, (these are summarized in Hutchinson et al. (1993)), they are placed in the two temporary test tables: M_DAILY_TEMP_TEST and M_HOURLY_TEMP_TEST. The structure of these two tables is identical to that of the permanent meteorologic tables, M_DAILY and M_HOURLY. Data in the temporary test tables comes directly from the daily and hourly temporary loading tables (M_HOURLY_TEMP_LOAD and M_DAILY_TEMP_LOAD) on a monthly basis. Quality control checks are performed on this data and some editing is done where necessary. These checks are described in Hutchinson et al. (1993). Upon completion of the quality assurance procedure, the records are appended to the permanent tables, M_DAILY and M_HOURLY.

M_HOURLY_TEMP_TEST_UNIQ_INDEX and M_DAILY_TEMP_TEST_UNIQ_INDEX are the unique indexes created for the hourly and daily temporary test tables.

4. DATABASE SUMMARY LISTING

A summary of all parameters collected at DRC is available to users. The summary indicates the percent real data by station by hydrologic year for each parameter in the M_DAILY table. Also included are hydrologic year means and totals for each DRC parameter.

The programme is called "met_summary" and may run from any directory on the HP-9000. Four listings are generated from this programmen ASCII format. These are called met_summary1.lst, met_summary2.lst, met_summary3.lst, met_summary4.lst and will be located in the current working directory of the user.

5. DATABASE LOADING AND UPDATE PROCEDURES

The two permanent data tables, M_DAILY and M_HOURLY, are updated (i.e. new information is added) on a regular schedule. Source data for these tables originates from AES and the DRC data loggers. As described in Section 3, to maintain data integrity of the tables, rigorous testing and quality control checks are run on the data prior to the update. Meteorologic database ancillary tables are updated as new collection stations are added or as methods change.

5.1 Database Loading

Dorset Research Centre (DRC) Data

DRC data loading can be divided into five steps; the retrieval of the data from the storage media (DRC data loggers), the formatting of the data for entry to the ORACLE database tables, the input of the data to the temporary loading tables, testing (checking) of the new data and finally, appending the "tested" data to the permanent tables. The following is a description of these steps.

i) **Data Retrieval**

Data from each of the four Limnology Section monitoring sites (PT1P,PCP2,HYP2 and HPP2), are transferred from the data logger onto a portable storage module each week. The module is then delivered to the DRC where the data are read from the module using the Campbell Scientific software (CSI 1989), "SMCOM". Data files created by this process are renamed conforming to predetermined standards and are stored on the met PC. Using the CSI software, "SPLIT", they are then reduced to hourly and daily files for each of the stations. Daily and hourly data is printed and visually checked for abnormalities and problems. Where problems are found they are flagged on the printouts, then filed into binders where they can be referenced at a later time (Data Integrity Testing).

Raw data files are transferred to the HP-9000 from the Met PC using the programme "MET_TRAN". This batch process uses the file transfer protocol "KERMIT" to move the data between the pc and mini computer across REFLECTION. The user remains logged on to the HP-9000 in the "METDATA" account after the file transfer and the raw data files reside in the /hydromet/metdata/raw directory where they can be referenced at a later time. This process is documented in Hutchinson et al. (1993).

ii) **Data Formatting**

The raw data files are generated in ASCII format and are comma delimited with one record per station date. A formatting procedure (met_format located in /usr/contrib/bin) is run on the transferred data to convert the numeric station code to a character station abbreviation (see Table 3). Met_format also changes the date format from "yy m d" to "yymmdd", as well it exchanges the hour value "2359" for "2400". The latter changes are required to insert the records into ORACLE.

Table 3 Station number/abbreviation information for DRC stations.

STN	Station # (logger)	STN id (LIS assigned)
PT1P	101	03 0085 536 23
HYP2	301	03 0085 506 23
PCP2	501	17 0021 541 23
HPP2	502	03 0085 571 23

iii) **Temporary Storage**

The data are loaded to the database temporary loading tables using the utility SQL*LOADER and two script files "load_dly_to_temp" and "load_hly_to_temp". An initial cleaning procedure (met_update1) is run on the tables to verify the existence of null values where parameters are not collected (e.g., due to data logger problems, some of the stations have erroneous values in the wind and radiation fields, these values must be set to null, until logger programme changes are made). Reference to M_PARAM_DAILY and M_PARAM_HOURLY for stations and parameters collected.

Data accumulates in the temporary tables for approximately four weeks at which time the database administrator moves the records to a set of testing tables where quality control checks are performed. Newly collected data continues to be loaded into the temporary tables.

iv) **Data Integrity Testing**

The quality of the data is evaluated in the temporary testing tables. The data is checked, and updated where necessary, against the notes (if any) made on the computer printouts (see Step 1. Data Retrieval). Further, more extensive, checking is done using the Met Data Management System (MDMS). MDMS is a menu driven programme which executes a series of SQL queries on the tables. The query results are spooled to listing files for later examination. Based on the listing contents, the user can make informed decisions regarding

changes to the data. Hutchinson et al. (1993) describes the MDMS programme with notes on its operation.

The first checks to be run on the tables are QC checks which examine station information, date values, data completeness and data value ranges. Updates are made to questionable data and the QC checks are run again. This is repeated until the listing produced by the checks is free of problem data.

A second set of checks made to the data are comparisons between daily and hourly table figures. Daily figures on a given station and day should be approximately equivalent to the sum or mean (parameter dependant) of the hourly values of the same station and date. Data in which daily values are not equivalent to hourly totals (within a specified tolerance range) are recorded in a listing file. Refer to Hutchinson et al. (1993) for a detailed description of all checking and editing procedures.

Prior to making corrections to this data, a third check should be run on the tables. This check identifies (by station and date) those records which are missing from the database. A complete record set in the daily table is comprised of one record per calendar date per station, and for the hourly table one record per hour per calendar date per station. Missing records found by this check may account for discrepancies between the totals in the hourly and daily data.

Two further important checks made by MDMS are for precipitation and wind. The precipitation check programme (PCHECK) compares precipitation depth data in the meteorologic database against precipitation depth and volume data in the WATER table. The depth data in the WATER table is computed by field samplers and entered to the database along with precipitation chemistry. There are frequent discrepancies between data logged values and field crew computations. The WIND programme checks direction and magnitude of the daily wind vector against the sum of the directions and magnitudes of hourly wind vectors.

The execution of these checks provide listings of problematic data values which are carefully compared against raw data files, printouts, field books, and Belfort charts. Care is taken to preserve the original information content of the measurements.Values are only removed when a data logger malfunction can be proved, or when the values exceed the normally accepted ranges listed in Hutchinson et al. (1993).

v) **Appending**

Upon completion of the quality control checks to the new data, the records may be appended to the permanent tables M_HOURLY and M_DAILY. The database administrator is responsible for this procedure and as a safety measure, must first archive the permanent tables by exporting them to digital audio tape (DAT).

The number of records in both the permanent and testing tables are then counted and recorded in the ORACLE log book in the computer room. Next, the records are appended into the permanent tables and a final table count is determined and logged to verify the addition of the new records. The testing tables are then dropped from the database.

Final checks are made to the data after they have been loaded into the permanent tables. The same checks are run over the permanent tables as were run over the temporary testing tables. Any data values requiring changes will be updated by the programme UPDATE. Using this programme both the old and new value are added to the M_UPDATE table and the old value is replaced in the permanent data table.

ATMOSPHERIC ENVIRONMENT SERVICES (AES) DATA

Availability of current AES data has a typical lag time of 6 to 18 months. This is due to collation of field records to digital data files, the quality assurance procedures run at the station monitoring site, and again at the Downsview computer centre. The station information, date values and data ranges are checked before the information is available.

Further quality control checks are not currently run on the AES data by the Dorset Research Centre thus, it is considered uncorrected data.

The update of AES data to the database is also a four step process; the request for data, the retrieval of data from media, formatting of data and data loading into ORACLE. The following are brief explanations of these steps.

i) **Request for Data**

AES data are formally requested every 6 months. This requested data is stored on 9 track magnetic 1600 bpi tape in defined record and block length files. Data are requested to cover a hydrologic year period (June 1 to May 31) to correspond to primary data use (water budgets for the calibrated study watersheds). A sample data request letter and complete Climatological Station catalogue are contained in Hutchinson et al. (1993).

ii) **Data Retrieval**

All data retrieval procedures are documented in Hutchinson et al. (1993). The data is read from the magnetic tape onto the HP-9000 using a UNIX script called "magtaperead". This script reads the data into ASCII format using the unix command "dd" and then parses the data into specified record lengths using either "PARSER232" (232 byte daily records) or "PARSER185" (185 byte hourly records).

iii) **Data Formatting**

The ORACLE database requires that each record added must contain all information for a given sample date. These files must therefore be separated by station, and then all measured elements for a sample date be joined (see ORACLE tables M_PARAM_DAILY and M_PARAM_HOURLY for elements collected at each station). The HP-9000 statistics package STATIT is used. Daily data files can be parsed to station files using DRC application procedures "ATMOS1" and "ATMOS2". These procedures are documented in Hutchinson et al. (1993).

iv) **Data Loading**

The database administrator uses SQL*LOADER to load the AES data directly into the permanent tables where it becomes available to users.

5.2 Database Back-up Strategy

Monday through Friday all ORACLE tables on the HP-9000 are exported into one file, (/oracle/"sysdate".dmp, where "sysdate" is the date on which the export was performed). Following the export each evening, a full system back-up is executed to DAT (Digital Audio Tape). All back-ups are stored in a locked fire-proof safe. The room containing the safe is only accessible during normal business hours. ORACLE tables can be restored from these archives by first restoring the exported file ("sysdate".dmp) to the system and then importing the table.

6. REFERENCES

AES. 1991. MANOBS, Manual of Surface Weather Observations., Seventh Edition. Atmospheric Environment Services, Canada.

Campbell Scientific Inc. 1989. PC208 Data logger support software instruction manual.

Futter, M.N., B.A. Hutchinson, A. Morgan and L.D. Scott. 1993. Hydrologic database storage for the Dorset Research Centre. Ont. Min. Envir. Data Report DR 93/3.

Hutchinson, B.A., M.N. Futter, A. Morgan and W. Coté 1993. Meteorologic data collection for the Dorset Research Centre. Ont. Min. Envir. Internal Data Report.

LaZerte, 1990 Q Documentation. Ont. Min. Envir. Energ. Internal Data Report.

LaZerte, B.D., D. Gardner and S. McCormick. 1990. Implementation of the database WATER at the Dorset Research Centre. Ont. Min. Envir. Internal Data Report.

Locke, B.A. and L.D. Scott. 1986. Studies of lakes and watersheds in Muskoka-Haliburton, Ontario: methodology (1976-1991). Ont. Min. Envir. Data Report DR 86/4.

Locke, B.A. and E. deGrosbois. 1986. Meteorological database for the Muskoka-Haliburton area. Ont. Min. Envir. Data Report DR 86/5.

ORACLE Corp. 1990a. SQL*Net TCP/IP Documentation.

ORACLE Corp. 1990b. SQL*Net NMP Named Pipes Documentation.

ORACLE Corp, 1991 - ORACLE System Documentation.

Pawson, T.W. and N.D. Yan. 1993. Zooplankton data and sample archive database design for the Dorset Research Centre. Ont. Min. Envir. Tech. Report. 38 pp.

APPENDIX 1

A. This query selects the average temperature, relative humidity and the number of stations reporting those parameters on a given date for the four meteorologic sites monitored by the Dorset Research Centre. It also selects the total radiation recorded at the Paint Lake site.

SQL*Plus>

```
1   select a.sdate,avg(a.avg_temp)avg_temp,count(a.avg_temp)t_num,
2   avg(a.avg_rh)avg_rh,count(a.avg_rh)rh_num,b.total_rad
3   from m_daily a,m_daily b
4   where a.stn in ('PCP2','PT1P','HPP2','HYP2')
5   and b.stn = 'PT1P'
6   and a.sdate = b.sdate
7   and a.sdate between '01-Jun-90' and '10-Jun-90'
8*  group by a.sdate,b.total_rad
```

SDATE	AVG_TEMP	T_NUM	AVG_RH	RH_NUM	TOTAL_RAD
01-JUN-90	15.7	3	55.3	3	28162
02-JUN-90	19.6	3	74.3	3	14891
03-JUN-90	17.6	3	75.8	3	16641
04-JUN-90	8.7	4	70.0	4	22492
05-JUN-90	10.7	4	63.6	4	28237
06-JUN-90	14.1	4	77.4	4	16320
07-JUN-90	15.4	4	67.5	4	29665
08-JUN-90	14.2	4	77.9	4	7547.7
09-JUN-90	15.6	4	83.0	3	13872
10-JUN-90	13.8	3	83.1	3	10301

B. This query selects the maximum and minimum temperatures recorded at a given month in 1988 from the Plastic Lake and Harp Lake meteorologic sites. Note that the Paint Lake site was not operational for two months in the winter of 1988.

SQL*Plus>

```
1   select to_char(a.sdate,'YYMM')yymm
2   min(a.min_temp)hp_min,max(a.max_temp)hp_max
3   min(b.min_temp)pc_min,max(b.max_temp)pc_max
4   from m_daily a, m_daily b
5   where a.stn = 'HPP2'
6   and b.stn = 'PCP2'
7   and to_char(a.sdate,'YYMM') = to_char(b.sdate,'YYMM')
8   and a.sdate between '01-Jan-88' and '31-Dec-88'
9*  group by to_char(a.sdate,'YYMM')
```

YYMM	HP_MIN	HP_MAX	PC_MIN	PC_MAX
8801	-36.5	6.3	-28.0	-2.3
8802	-29.3	6.8	-34.5	6.4
8803	-27.5	14.1		
8804	-7.8	18.1		
8805	-1.7	32.7	0.7	31.1
8806	-2.4	34.4	1.0	34.2
8807	5.6	37.9	8.1	35.3
8808	3.3	36.1	6.6	36.1
8809	-0.4	26.5	1.9	26.4
8810	-5.5	23.0	-0.8	22.3
8811	-10.1	15.7	-5.0	12.5
8812	-29.8	7.3	-27.3	6.7